Concurrent Acquisition of a Single Nucleotide Polymorphism in Diverse Influenza H5N1 Clade 2.2 Sub-clades

Niman HL*, and Boynton BR+

Recombinomics, Inc., Pittsburgh, Pennsylvania, USA*, U.S. Naval Medical Research Unit 3 (NAMRU-3), Cairo, EGY+

Department of Influenza Recombination, Recombinomics, Inc, 648 Field Club Road, Pittsburgh, PA 15238, USA

Correspondence to Henry L Niman, E-mail: henry_niman@recombinomics.com

617.877.0987, 412.963.1362 FAX

Text word count: 1691

3 Figures 1 Supplemental Table

Running Title: Concurrent Influenza H5N1 Acquisitions

Highly pathogenic Influenza A H5N1 was first identified in Guangdong Province in 1996, followed by human cases in Hong Kong in 1997 [1,2]. The number of confirmed human cases now exceeds 300, and the associated Case Fatality Rate exceeds 60% [3]. The genetic diversity of the serotype continues to increase. Four distinct clades or sub-clades have been linked to human cases [4-7]. The gradual genetic changes identified in the sub-clades have been attributed to copy errors by viral encoded polymerases that lack an editing function, thereby resulting in antigenic drift [8]. We report here the concurrent acquisition of the same polymorphism by multiple, genetically distinct, clade 2.2 sub-clades in Egypt, Russia, and Ghana. These changes are not easily explained by the current theory of “random mutation” through copy error, and are more easily explained by recombination with a common source. This conclusion is supported by additional polymorphisms shared by clade 2.2 isolates in Egypt and Germany.

The study of influenza evolution in nature has been aided by the emergence of a new strain (clade 2.2) first identified at Qinghai Lake in central China in the spring of 2005. Sequencing of all eight genes [9,10] showed that isolates from migratory waterfowl were easily distinguishable from previous isolates linked to poultry and human infections in eastern and southeastern Asia [11,12]. The new strain was subsequently found in outbreaks in Russia, Kazakhstan, and Mongolia [13,14]. Prior to these clade 2.2 outbreaks, the highly pathogenic Asian version of H5N1 had never been reported west of China. The detection of H5N1 in migratory waterfowl in the summer of 2005 in migratory bird sanctuaries in Russia and Mongolia signaled the start of a major geographical expansion of H5N1. In the following 12

months, almost 50 countries west of China reported H5N1 for the first time, and all infections were clade 2.2. The geographical reach included Europe, the Middle East, and Africa.

This expansion offered a unique opportunity to study H5N1 evolution linked to migrated into new regions. Infections included human cases in Turkey, Iraq, Azerbaijan, Egypt, Djibouti in 2006 Nigeria in 2007. Sequence analysis indicated all cases were due to clade 2.2. The outbreaks were due to multiple introductions, and isolates had region specific polymorphisms.

The sequences also allowed for the monitoring of new acquisitions of discriminating polymorphisms. These new acquisitions created additional sub-lades defined by phylogenetic analysis. Moreover, close monitoring of these changes provided insight into the mechanisms underlying the rapid evolution of H5N1 in general, and sub-clade 2.2 in particular.

We isolated H5N1 from patients and poultry in Egypt. The first poultry isolates were collected February, 2006. The first human cases developed symptoms in March, 2006. Analysis of the H5N1 isolates collected between February and May. 2006 defined a series of HA and NA regional markers. These markers were also found in the human case from Djibouti, as well as in poultry isolates from Israel and Gaza. These isolates are listed in supplemental table 1.

After a lull in reported infections over the summer, H5N1 re-emerged in Egypt in September, 2006. The more recent isolates had the same regional markers seen in the previous season. However, both poultry and human isolates had acquired a series of new polymorphisms. Non-synonymous polymorphisms were identified in samples collected from a cluster of three family members from the Gharbiya governorate in the Nile Delta. HA gene polymorphisms were identified in or near the receptor binding domain, including V223I and M230I, as well as the oseltamivir resistance polymorphism N294S, in the NA gene. The patients first developed symptoms in December 2006 and all three infections were fatal (a detailed report on patients and polymorphism tracing will be described elsewhere).

Additional cases in early 2007 included HA sequences with a 3 BP deletion of the nucleotides encoding Ser at position 133 (H3 numbering), as well as a case with a novel HA cleavage site, REGRRRKR. The changes were found in multiple patients in central and southern Egypt. The above non-synonymous changes were associated with additional synonymous and non-synonymous changes in the HA and NA sequences that created additional sub-clades of subclade 2.2. However, the isolates from the 2006/2007 season maintained the regional markers seen in early 2006 in isolates from Egypt, Djibouti, Israel, and Gaza.

Chicken isolates from Gharbiya samples collected on February 15, 2007 included one sequence that was closely related to the sequences from the human Gharbiya cluster. The sequence from this isolate, A/chicken/1892N3-HK49/2007

(HK49), had the regional markers previously seen in the 2006 and 2007 isolates, as well as HA non-synonymous changes, V223I and M230I. Additionally, it also had an NA synonymous polymorphism, G743A, appended onto the genetic background of the human Gharbiya cluster, as seen in the NA cladogram in Figure 1. The location of the isolate on the tip of the branch indicated it was a recent acquisition.

This polymorphism was also found in two additional chicken isolates, A/chicken/1890N3-HK45/2007 and A/chicken/1891N3-CLEVB/2007 collected the same week in the Gharbiya governorate, but these two isolates fell onto a separate branch of the tree, indicating the same polymorphism had been appended onto two distinct sub-clades. Moreover, plaque-purified sub-clones of HK49 were isolated because the original sequence had mixed signals in the NA and HA sequences. These HK 49 sub-clones fell into two major groupings. The NA consensus sequences for the two groups had 11 differences, which matched the 11 differences between the two sets of chicken sequences. The major species were closely related to the sequences from the human cluster, while the minor species were closely related to the two additional chicken sequences. However, all sequenced plaque purified clones had G743A (data not shown).

The G743A was subsequently found in human isolates from patients who developed symptoms in April, 2007. Included were siblings with HA sequences that had the 3 BP deletion seen in earlier patients from central Egypt. Like the chicken sequences above, the G743A polymorphism was appended onto

sequences identified earlier in Egypt. Similarly, distinct sequences from another patient, A/Egypt/2630-NAMRU3/2007, which also acquired G743A, fell onto a separate branch.

The distinct branches displayed in the NA cladogram were also seen in the HA cladogram in Figure 1B. The isolates with G743A are also located at the tips of the branches, supporting a recent acquisition of the polymorphism.

In February, 2007, an H5N1 clade 2.2 outbreak occurred near Moscow, Russia. Isolates from infected chickens were most closely related to 2006 sequences from Azerbaijan. Figure 2A is an expanded cladogram with isolates from Europe, the Middle East, and western Africa. Like the acquisitions in Egypt, the isolates with G743A mapped onto the tip of a branch composed of earlier isolates that did not have the acquisition.

Similarly, in April, 2007, an H5N1 clade 2.2 outbreak occurred near Tema, Ghana. Sequences from three chickens were most closely related to turkey isolates collected in December, 2006 in the Ivory Coast. Like the Egypt and Moscow isolates above, the Ghana sequences with G743A mapped to the tip of a branch containing earlier isolates that did not have that polymorphism.

Additional HA polymorphisms are noted in the HA phylogram in Figure 2B. Isolates that had the NA polymorphism, G743A, also had a synonymous HA polymorphism, C689T. This polymorphism was also in human and bird isolates from the Nile Delta (see supplemental table 1). Another polymorphism, G754A,

that encodes M230I is in one of the German isolates [15], A/eagle owl/Germany/R166/2006, and maps to another branch with Egyptian human and poultry isolates from the Nile Delta. A third polymorphism, C1614T, that encodes T517I, is in another German isolate, A/mute swan/Germany/R797/2006, and on another branch with human isolates from southern Egypt. The isolates also have the novel HA cleavage site initially found in whooper swan isolates in Mongolia in 2005. The polymorphisms found in German isolates in 2006 were in Russian clade 2.2 isolates in 2005.

The NA G743A polymorphism can be traced through public H5N1 sequences. Isolates with full NA sequences are listed in the NA phylogram in Figure 3. The polymorphism was identified in the first reported sequences linked to the spread of H5N1 in Asia in 2003/2004 in South Korea [16] and Japan [17]. The polymorphism was subsequently identified in clade 1 isolates in southeast Asia, as well as clade 2.1 isolates in Indonesia and clade 2.3 isolates in China. The first reported clade 2.2 isolates were in wild birds in Germany (see figure 2) collected in February, 2006. The isolates in Germany formed distinct HA and NA branches due to a series of regional markers in these isolates.

The concurrent acquisition of the same polymorphism by multiple sub-clades challenges the current theory of influenza evolution that invokes random mutations as a mechanism for the generation of antigenic drift. The isolates with the newly acquired polymorphisms map to the tips of the branches of the phylogenetic trees, indicating the acquisitions were recent. All discussed isolates

on the tips of the branches were collected over a short time frame between February and April, but in three geographically distinct regions. These data do not support a common progenitor sequence, because the most closely related sequences to each of the respective recent isolates do not have this change. Similarly, concurrent mutation / selection by the eleven isolates which map to six branches in three countries and collected over a short time frame is also unlikely.

An alternative explanation for the appending of these sequences is through homologous recombination between closely related sequences. The newly acquired polymorphisms in the recent isolates in Egypt are readily found in recent H5N1 isolates, as noted in the three HA examples previously described above. Moreover, the G743A polymorphism is found in these genetically distinct sequences collected over short time frame.

Polymorphism tracing demonstrates that most of the newly acquired polymorphisms can be traced to the same serotype identified recently at locations that are linked together by migratory bird flyways, raising the possibility that the distribution and acquisition of the polymorphism is linked to recombination between H5N1 sequences transported and transmitted by migratory birds.

The individual polymorphisms recombine, generating sequences that create antigenic drift. Mapping of these pathways and associations may be used to develop novel vaccine targets representing rapidly evolving genomes.

Figure 1 NA and HA Phylograms of Egyptian Isolates

A. NA phylogram of positions 43-1337. Isolates with G743A marked with red arrows. NA regional markers are C150T, C236T, A703G, A740G, T1088C, G1280A. Accession numbers and additional isolates with G743A listed in table S1.

B. HA phylogram of positions 93-1688. Isolates with G743A marked with red arrows. HA regional markers are G196A, G466A, G496A, C661T, C727T, C779T, A878G, C937T, G1018T, C1261T, C1686T.

Egyptian isolates, accession numbers, collection date, and governorate in Table S1. Other public sequences with accession numbers in Table S2. Trees generated using neighbor joining with 100 bootstrap repetitions. Sequences generated as described previously [18].

Figure 2 NA and HA phylogram of Clade 2.2 isolates.

A. NA phylogram of positions 43-1337. Isolates with G743A marked with red arrows. Accession numbers and additional isolates with G743A listed in table S1.

B. HA phylogram of positions 93-1688. Isolates with C628T marked with blue arrows / bars. Isolates with G754A marked with green arrows / bars. Isolates with C1614T marker with orange arrows / bars.

Accession numbers and additional isolates with HA polymorphisms listed in table S1

Figure 3 NA Phylogram of H5N1 isolates with G743A

Accession numbers and additional isolates with G743A listed in table S1.

Table 1S Isolates and Accession Numbers

Names and accession numbers of HA and NA sequences used in figures 1-3. Partial sequences [19-26] with polymorphisms in Figures 1-3 are listed.

Acknowledgements

We thank Magdi Sadd, Jeff Tjaden, Ken Earhart, Marshall Monteville, Moustafa Mansour, Elassai Labib, EA Ayoub from the US Naval Medical Research Unit 3, Mona Aly and A-SA Arasa from the Central Laboratory from Veterinary Quality Control, and Al Nayel, AS Abdelghani, Hala Essmat from Ministry of Health, Arab Republic, for the collection of human and animal samples, and isolation and sequencing of H5N1.

References

1. Subbarao, K. *et al*. Characterization of an Avian Influenza A (H5N1) Virus Isolated from a Child with a Fatal Respiratory Illness *Science* **279**, 393-396 (1998).

2. Claas, E.C. *et al.* Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet* **351,** 472-477 (1998).

3. Cumulative number of confirmed cases of human influenza A/(H5N1) reported to WHO. Available from http://www.who.int/csr/disease/avian_influenza/country/cases_table_2007_05_16/en/index.html

4. Peiris, J.S. *et al*. Re-emergence of fatal human influenza A subtype H5N1 disease. *Lancet* **363**, 617-619 (2004).

5. Fouchier,R., Kuiken, T., Rimmelzwaan, G., & Osterhaus, A. Global task force for influenza.*Nature* **435**, 419-420 (2005).

6. Osterholm, M.T. Preparing for the next pandemic. *N. Engl. J. Med*.**352,** 1839 - 1842 (2005).

7. Monto, A.S. The Threat of an Avian Influenza Pandemic. *N. Engl. J. Med*. **352**, 323 - 325 (2005).

8. Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M. & Kawaoka, Y. Evolution and ecology of influenza A viruses *Microbiol. Rev*.**56**, 152 - 179 (1992).

9. Chen, H. *et al*. Avian flu: H5N1 virus outbreak in migratory waterfowl. *Nature* **436**, 191-192 (2005).

10. Liu, J. *et al*. Highly Pathogenic H5N1 Influenza Virus Infection in Migratory Birds *Science*, **309**, 1266-1267 (2005).

11. Li, K.S. *et al*. Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature*. **430,** 209-213 (2004).

12. Guan, Y. *et al*. H5N1 influenza: A protean pandemic threat. *Proc. Natl. Acad. Sci. U.S.A.***101,** 8156 – 8161 (2004).

13. Shestopalov, A.M. *et al*. H5N1 Influenza Virus, Domestic Birds, Western Siberia, Russia. *Emerg. Infect. Dis*. **12**, 1167-1169 (2006).

14. Aleksandr, S. *et al*. Influenza (H5N1) Viruses in Poultry, Russian Federation, 2005–2006. *Emerg. Infect. Dis*. **13**, April (2007). Available from http://www.cdc.gov/eid/content/13/4/539.htm

15. Weber, S. *et al*. Molecular analysis of highly pathogenic avian influenza virus of subtype H5N1 isolated from wild birds and mammals in northern Germany. *J. Gen. Virol.* **88**, 554-558 (2007).

16. Lee, C.W. *et al.* Characterization of Highly Pathogenic H5N1 Avian Influenza A Viruses Isolated from South Korea. *J. Virol*. **79**, 3692-3702 (2005).

17. Mase, M. *et al.* Characterization of H5N1 influenza A viruses isolated during the 2003-2004 influenza outbreaks in Japan. *Emerg. Infect. Dis.* **11**, 699-701 (2005).

18. Yingst, S.L., Saad, M.D., & Felt, S.A. Qinghai-like H5N1 from domestic cats, Northern Iraq. *Emerg. Infect. Dis.* **12**, 1295-1297 (2006).

19. Ducatez, M.F. *et al.* Avian flu: multiple introductions of H5N1 in Nigeria. Nature **442**, 37 (2006).

20. Salzberg, S.L. *et al*. Genome analysis linking recent European and African influenza (H5N1) viruses. *Emerg. Infect. Dis.***13,** May (2007). Available from http://www.cdc.gov/EID/content/13/5/713.htm

21. Ducatez, M.F. et al. Genetic Characterization of HPAI (H5N1) Viruses from Poultry and Wild Vultures, Burkina Faso. *Emerg. Infect. Dis.* **13,** April (2007). Available from http://www.cdc.gov/eid/content/13/4/611.htm

22. Thanawongnuwech, R. *et al*. Probable tiger-to-tiger transmission of avian influenza H5N1. *Emerg. Infect. Dis*. **11**, 699-701 (2005).

23. Smith, G.J. *et al*. Evolution and adaptation of H5N1 influenza virus in avian and human hosts in Indonesia and Vietnam. *Virology* **350**, 258-268 (2006).

24. Chen, H. *et al*. Establishment of multiple sublineages of H5N1 influenza virus in Asia: implications for pandemic control. *Proc. Natl. Acad. Sci. U S A.* **103**, 2845-2850 (2006).

25. Smith, G.J. *et al* Emergence and predominance of an H5N1 influenza variant in China. *Proc. Natl. Acad. Sci. U S A.* **103,** 16936-16941 (2006).

26. Payungporn, S, *et al.* Single step multiplex real-time RT-PCR for H5N1

influenza A virus detection. *J. Virol. Methods* **131**, 143-147 (2006)

Figure 1A NA Egypt G743A

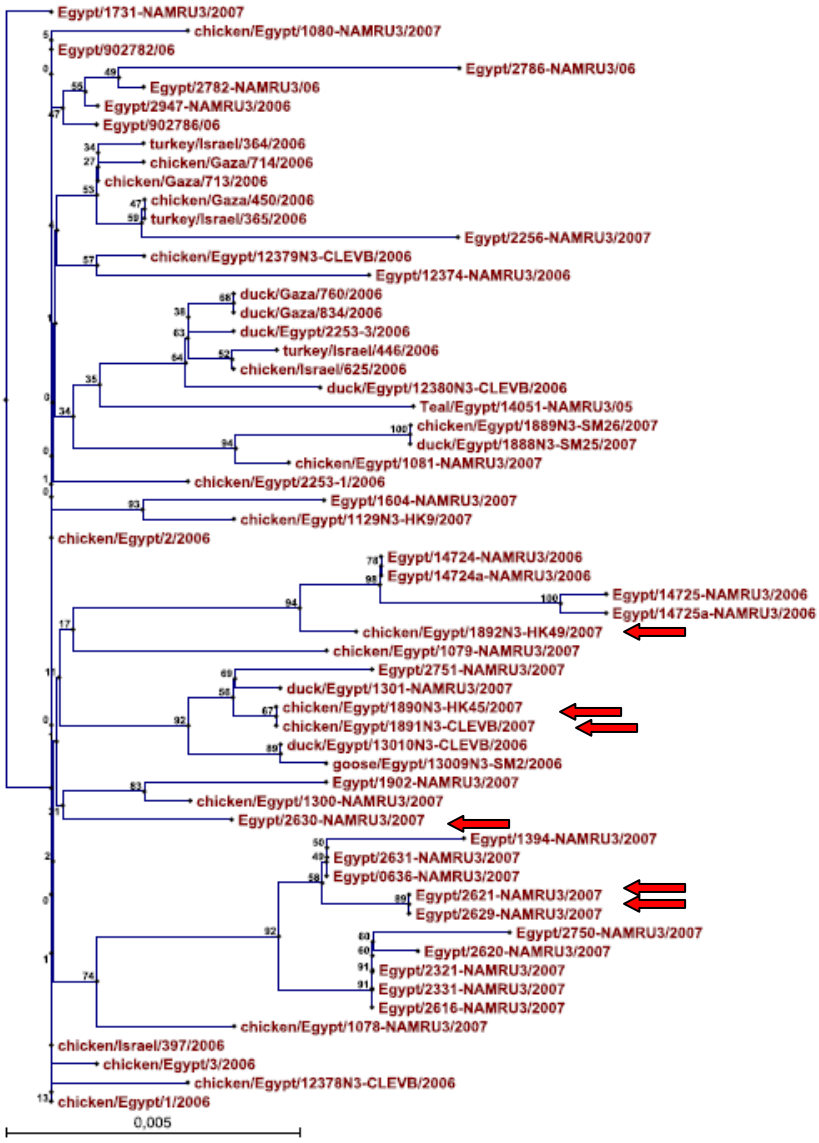


Figure 1B HA Egypt G743A

Figure 2A NA Phylogenetic Tree

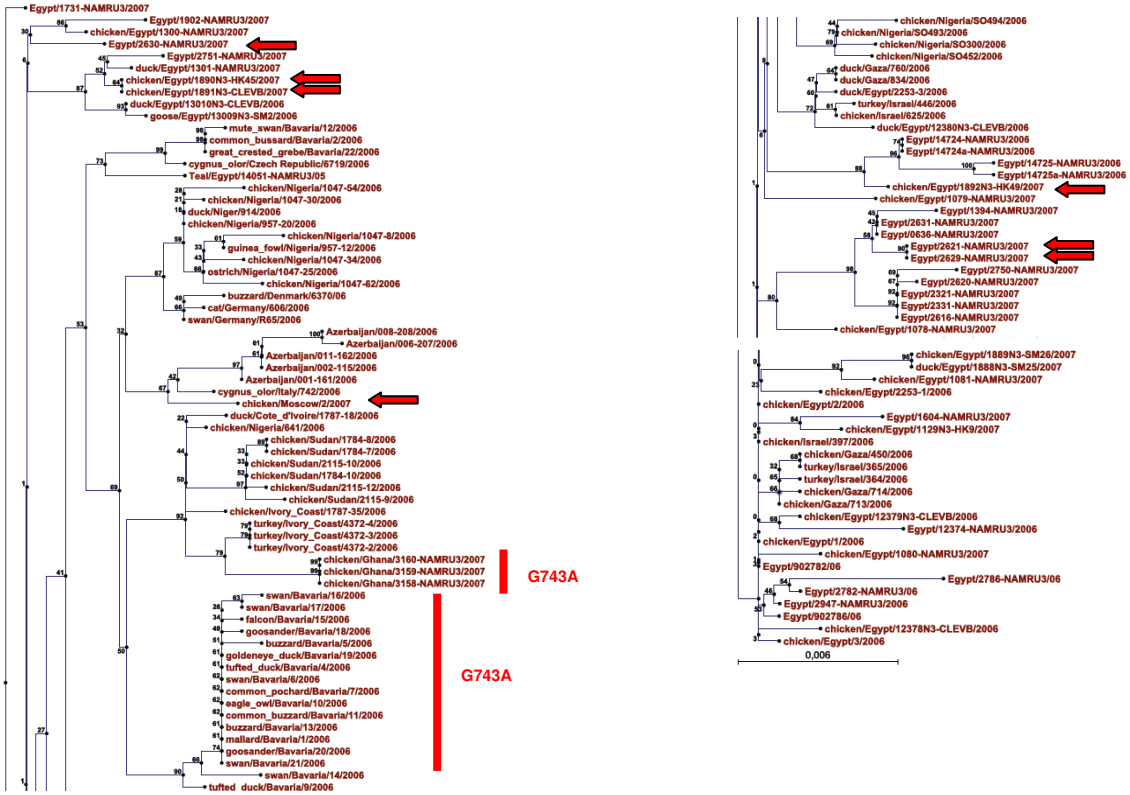
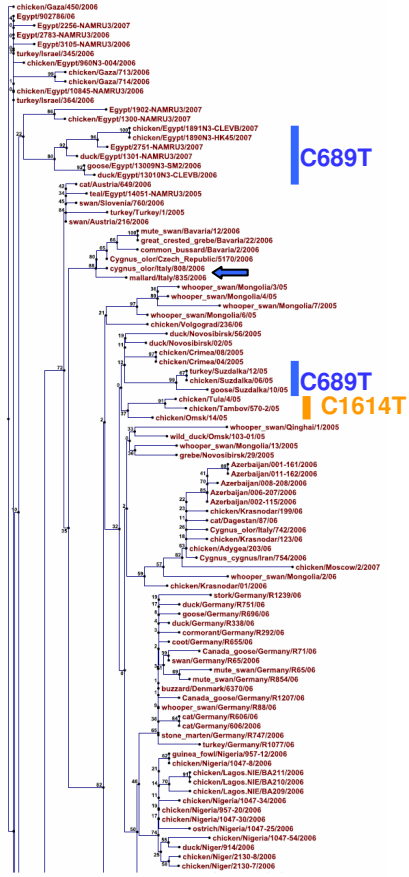


Figure 2B HA Phylogenetic Tree



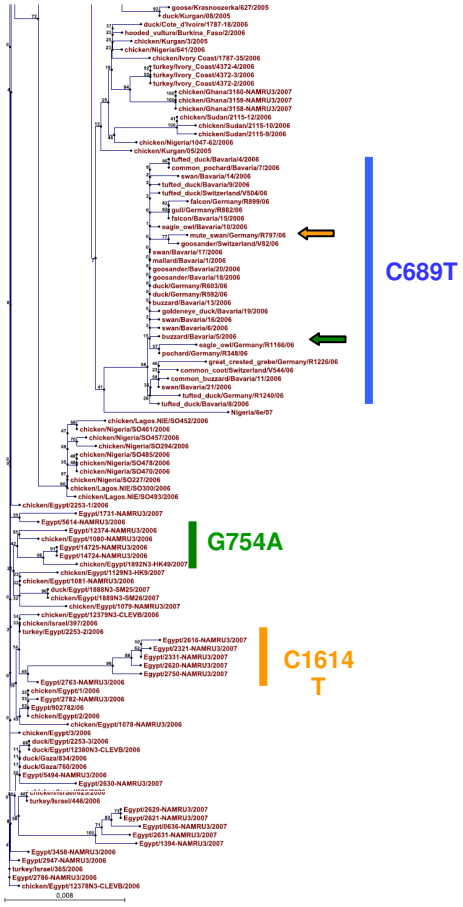


Figure 3 NA G743A